

CLAIMS

What is claimed is:

1. A method for generating hyperlinks, comprising:
locating a text reference in a source document;
5 identifying a target document relating to the text reference;
deriving an anchor text corresponding to the target document utilizing the
source document;
generating a hyperlink to the target document; and
associating the hyperlink with the anchor text.
- 10 2. The method of claim 1, wherein locating the text reference comprises
deriving the text reference based on a statistical model of at least one of text formatting
and lexical cues.
3. The method of claim 1, wherein locating the text reference comprises
comparing text from the source document with a list of predetermined references.
- 15 4. The method of claim 1, further comprising:
locating a label corresponding to the text reference; and
associating the hyperlink with the label.
5. The method of claim 4, wherein the locating the label comprises deriving
the label based on a statistical model of at least one of text formatting and lexical cues.

6. The method of claim 4, further comprising deriving a label anchor text depending on whether the label corresponding to the text reference precedes or follows a text phrase.

7. The method of claim 6, wherein the label anchor text is a longest noun phrase extracted from the text phrase following or preceding the label when the label precedes or follows the phrase, respectively.

8. The method of claim 1, further comprising parsing the text reference into a plurality pieces of text, wherein the identifying, deriving, generating, and automatically associating are performed for each of the plurality pieces of text.

9. The method of claim 1, wherein the source document is selected from the group consisting of an HTML document, a text document, a postscript document, a Portable Document Format (PDF) document, a PowerPoint document, a Word document, an Excel document, and a close-captioned video.

10. The method of claim 1, wherein the text reference is a reference to one of a paper, article, company, institution, product, search engine, image, object, and geographical location.

11. A system for generating hyperlinks, comprising:
- a text reference locator configured to locate a text reference in a source document;
 - a document identifier configured to identify a target document relating to the text reference;
 - an anchor text determining engine configured to compute an anchor text corresponding to the target document; and
 - a hyperlink generator configured to generate a hyperlink to the target document and to automatically associate the hyperlink with the anchor text.
12. The system of claim 11, wherein the text reference locator is further configured to locate the text reference based on a statistical model of at least one of text formatting and lexical cues.
13. The system of claim 11, wherein the text reference locator is further configured to locate a label corresponding to the text reference and wherein the hyperlink generator is further configured to associate the hyperlink with the label.
14. The system of claim 13, wherein the text reference locator is further configured to locate the label based on a statistical model of at least one of text formatting and lexical cues.

15. The system of claim 13, wherein the anchor text determining engine is further configured to determine a label anchor text depending on whether the label corresponding to the text reference precedes or follows a text phrase.

16. The system of claim 15, wherein the label anchor text is a longest noun phrase extracted from the text phrase following or preceding the label when the label precedes or follows the phrase, respectively.

17. The system of claim 11, wherein the text reference locator is further configured to parse the text reference into a plurality pieces of text, wherein the document identifier, anchor text determining engine, and hyperlink generator are executed for each of the plurality pieces of text.

18. The system of claim 11, wherein the source document is selected from the group consisting of an HTML document, a text document, a postscript document, a Portable Document Format (PDF) document, a PowerPoint document, a Word document, an Excel document, and a close-captioned video.

19. The system of claim 11, wherein the text reference is a reference to one of a paper, article, company, institution, product, search engine, image, object, and geographical location.

20. A computer program product embodied on a computer-readable medium, the computer program product including instructions, which when executed by a computer system, are operable to cause the computer system to perform acts comprising:

- 5 locating a text reference in a source document;
- identifying a target document relating to the text reference;
- deriving an anchor text corresponding to the target document utilizing the source document;
- generating a hyperlink to the target document; and
- associating the hyperlink with the computed anchor text of the text
- 10 reference.

21. The computer program product of claim 20, wherein the locating the text reference comprises computing the text reference based on a statistical model of at least one of text formatting and lexical cues.

22. The computer program product of claim 20, further including instructions
15 operable to cause the computer system to perform acts comprising:

- locating a label corresponding to the text reference; and
- associating the hyperlink with the label.

23. The computer program product of claim 22, wherein the locating of the label comprises computing the label based on a statistical model of at least one of text
20 formatting and lexical cues.

24. The computer program product of claim 22, further including instructions operable to cause the computer system to perform acts comprising:

computing a label anchor text depending on whether the label corresponding to the text reference precedes or follows a text phrase.

5 25. The computer program product of claim 24, wherein the label anchor text is a longest noun phrase extracted from the text phrase following or preceding the label when the label precedes or follows the phrase, respectively.

26. The computer program product of claim 20, further including instructions operable to cause the computer system to perform acts comprising parsing the text
10 reference into a plurality pieces of text, wherein the performing the search, computing the anchor text, generating the hyperlink, and associating the hyperlink are performed for each of the plurality pieces of text.

27. The computer program product of claim 20, wherein the source document is selected from the group consisting of an HTML document, a text document, a postscript
15 document, a Portable Document Format (PDF) document, a PowerPoint document, a Word document, an Excel document, and a close-captioned video.

28. The computer program product of claim 20, wherein the text reference is a reference to one of a paper, article, company, institution, product, search engine, image, object, and geographical location.